Using Educational Data Mining to Identify and Analyze Student Learning Strategies

Gove Allen, Brigham Young University, gove@byu.edu
Randy Davis, Brigham Young University, randy.davies@byu.edu
Nicholas Ball, Utah Valley University, nick.ball.phd@gmail.com
Conan Albrecht, Brigham Young University, doconix@gmail.com
Nesrin Bakir, West Texas A&M University, nbakir@wtamu.edu

Abstract

This paper explores student learning strategies in an introductory spreadsheets course. Student study habits were tracked at a level of detail not available in previous research. Detailed data were collected regarding reading, video watching, actions in practice assignments, references to assignment instructions, and actions in graded assignments. The analysis indicates that student strategies cluster into four primary learning groups. The study provides insight into how instructors can develop their courses and lectures in ways that better match the learning strategies of their students.

1. Introduction

Helping students learn is the goal of every instructional system. However, instructional designers cannot guarantee learning will occur simply because students are given quality instructional materials. A researcher's ability to establish generalizable best practices and improve instructional systems is hindered by the human aspect of the social sciences, and the inability of researchers to control for all the variables likely to influence whether the desired learning outcomes of a course will be accomplished. The degree to which students learn is influenced by both personal and contextual factors. One often-studied factor related to learning includes the strategies students use to accomplish learning objectives in a course -- a decidedly meta-cognitive aspect of learning [1].

Learning strategies refer to the specific actions taken by the learner to accomplish the learning required of them in an educational situation [2]. Some strategies are more likely to produce good results than others; understanding how students go about the learning activities of a course can also help instructional designers improve their course in general, as well as making a course more adaptive when attempting to meet the individual needs of specific learners.

Research into learning strategies is not new; indeed, it has been studied for decades. However, previous research has been hindered by a significant disadvantage: studies in traditional classrooms had to collect data in using self-report questionnaires as a primary data source [3].

Collecting data via self-reporting can be a difficult process to control and validate. It places limits on the amount of data that can be collected because of time constraints and human errors. It must be done in ways that do not interrupt the natural processes being studied.

Advances in technology-enabled instruction have changed this limitation. Online environments are able to collect significant data in unobtrusive ways. However, new research is needed to better understand the strategies student use to complete courses delivered asynchronously online or in a blended class format.

In contrast with previous data collection methods, one challenge of studying student learning strategies in technology-enabled instructional systems is the overwhelming amount of data available [4]. While we no longer need to rely completely on student self-report to determine precisely what strategies students use when completing a course, deciding which data are important, capturing and linking all the relevant data, then creating actionable information from data can be a challenge.

The purpose of this study is to use an online environment to collect a large amount of data and use educational data mining techniques to better understand the learning strategies students utilize. In the study, we collect data from students completing a spreadsheet course. The course included video, reading, and assignments in Microsoft Excel.

URI: https://hdl.handle.net/10125/59442
ISBN: 978-0-9981331-2-6

In this paper, we attempt to discover and understand the strategies that students use when completing a technology-enabled, online course.

2. Previous Work

The research literature makes a distinction between learning approach and learning strategies. Leaning approaches are typically described as either deep learning or surface learning approaches [3]. Educators and researchers typically praise the virtues of deep learning and devise ways to encourage surface learners to engage more fully in the learning activities in order to learn all they can. Unfortunately, students do not always have the same academic goals as their instructors; they are often only intend to attain a sufficient level of learning to earn the grade they want [5]. A criticism of many course is that they are designed in such a way that deep learning is not rewarded and, in fact, not needed for students to pass a course; students can often achieve their learning goals with surface learning alone [6]. Through a meta-cognitive process, students devise learning strategies to accomplish their learning goals. These learning strategies may be intended to achieve either surface or deep learning.

Learning strategies students devise are based on personal factors including a student's academic goals, learning preferences, their self-efficacy and locus of control, as well as their ability to self-regulation [7]. Contextual factor that affect the learning strategies student choose include the difficulty of the task, a student's interest in the topic, as well as the affordances that the instructional design of the course provides to the students [8]. Strategies student use to accomplish instructional activities and tasks often reflect a student's desire to learn efficiently but often not always effectively [1]. There are many reasons for this; one reason being that students often have conflicting intentions – they have a lot of things to do and a limited time to do them [5]. Often a student will modify or change their learning strategies as the course progresses. The way a student approaches a learning situation is not inherent, it is developed by the learner and is often dependent on the learning context or situational demands [3]. Not all learning strategies are effective. Understand the strategies student use to complete courses can help educators and instructional designers improve their course and often provide actionable information that informs how and in what ways an educator might remediate learning gaps and students' misconceptions [9].

Research involving learning approaches and learning strategies have in the past relied primarily on self-report instruments [10]. In these studies, detailed records of topic focus, media choice, and study times and durations were difficult to collect. For example, understanding the strategies students use to complete an assignment might require collecting time spent on each problem, where and when students referenced their textbooks, and how students progress from initial answers to submitted answers. These data have been difficult to collect in reliable, efficient ways.

With advances in technology and increases in technology-enabled instruction, researchers are able to gather considerably more information about the strategies students use to complete the learning activities required for a course [4]. Capturing data within the system allows researchers to analyze the temporal order of spontaneous individual activities of students as they complete a course [11]. Not only does this allow researchers to obtain a more accurate description of students learning strategies, it can be the basis for real time implementation of adaptive practices.

3. Experiment

The subjects for this study are all students in an introduction to information systems course. The course covers a number of topics. The class consists of both lecture and hands-on computer lab sessions. During the lab sessions, students are exposed to topics in Microsoft Excel. About a third of the topics in the course are specific to students mastering elements of Microsoft Excel. This study focuses on student strategies and methods for learning Excel.

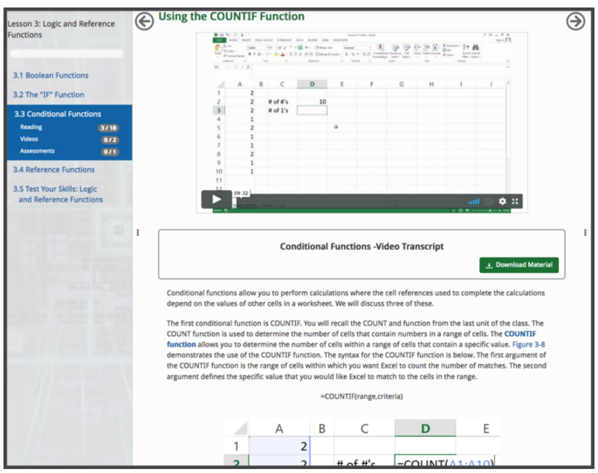
The students in the course are all undergraduate business students and are required to take the class. All of the students have basic computing skills (Internet, word processing, and email). Though the course does not require students to have prior experience with Microsoft Excel, some student enter the course with some familiarity with the application.

During the course, students work in two environments. The first is the MyEducator website, which hosts the textbook and videos. This website is shown in Figure 1. The website includes a "reader" that presents the textbook similar to normal introductory textbook: with chapters and sections, key terms and glossary. Students read the textbook on their laptops and mobile devices, and they can listen to the text similar to a podcast. The text can be highlighted with different "markers", and notes can be taken in the text.

The platform also provides some additional features not available in traditional textbooks. The left-side toolbar shows an outline of the current chapter, with topics that students have read marked with check mark icons. Learning tools like flashcards for key terms are available.

Each section of the text includes one or more video presentations by the author. Videos are embedded within each web page alongside the text, making access to both equally easy. The video content complements the text: students can choose to read, to watch video, or to do both. Videos had to be clicked by the student to play.

Figure 1: Textbook reader and video player



We have informally discussed the text and video content with students completing the course. Students self report high levels of use and satisfaction with the video elements of the instructional materials. Students comment positively about the convenience of watching the videos on demand and the fact that they can pause, rewind, and even watch the videos at an increased speed.

This anecdotal evidence suggests that the videos are a well-used and well-received element of the instructional materials. As we will show in this paper, this anecdotal evidence does not reflect the empirical usage patterns we observe with video use. The gap we see in how students report on their use of the instructional materials calls into question and veracity of self-reported data on student learning patterns.

In additional to the online text and video learning materials, several practice assignments are available in each chapter (generally one for each section within the chapters). Students are not required to complete practice assignments. Instead, students are encouraged to complete the practice problems to gain experience with the topics of each section. Students receive instantaneous feedback on their performance on the practice problems from an automated scoring and feedback system. Step-by-step instructions and a video are available for students to assist them in completing all practice problems.

Each chapter is matched with an assignment in Microsoft Excel that students complete for grades in the course. As with the practice problems, download the Excel files and complete them on their local computer. Students are also provided with instantaneous feedback on their performance on the assessments. In contrast to the practice problems, students are not provided with step-by-step instructions or a video to aid them in completing the chapter assessments. A detailed description of how students interact with the assignment files is presented in the data collection section of this paper.

Data collection

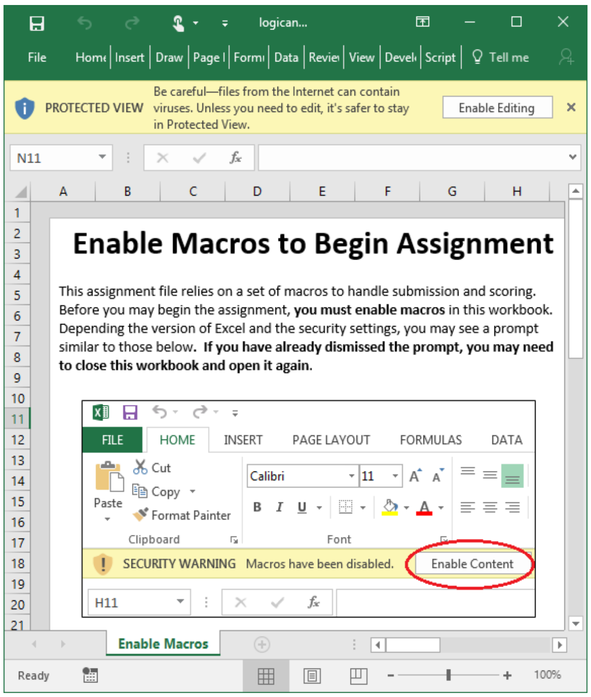
Data were collected on student actions in the textbook reader (website) and actions within the Excel workbooks. A total of 997 students were included in that analysis. These students completed all the lessons and assignments required in the course. The system captured student behavior in five categories: reading, video watching, practice assignments, primary assignments, and task guide views.

Student reading was tracked by client-side scripts that updated the server every 15 seconds and during page unload. As students read the textbook, they scrolled the browser window downward through the text. Whenever scrolling paused long enough, the paragraphs in view were deemed "read" by the student.

Embedded videos were split into 5-second blocks and tracked by block. As students watched (or skipped around in a video), the blocks that played were recorded as "seen".

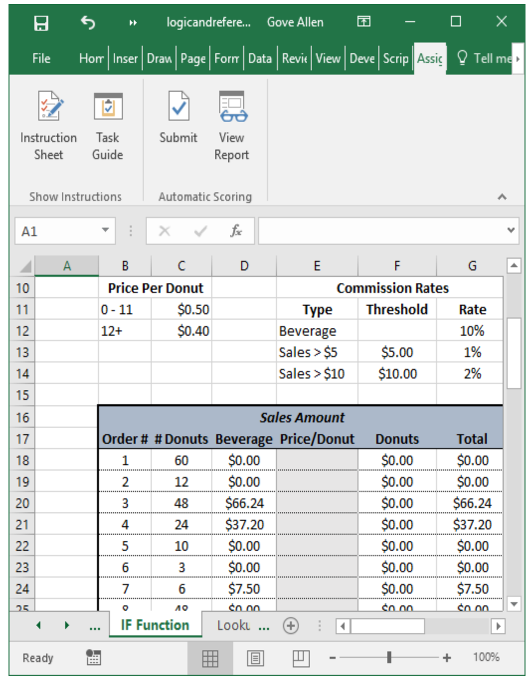
A student begins an assignment by downloading an Excel workbook from the MyEducator website. Using Visual Basic for Applications (VBA), the programming language built into Excel, the workbook will keep track of the student's progress as he or she completes the assignment and interact with the MyEducator servers during submission. Because of this, students must enable macros when the file is first opened. In fact, they the worksheets needed to complete the assignment are not made visible until the student does so. Figure 2 shows what the student sees upon opening an assignment workbook before enabling macros.

Figure 2. Initial workbook



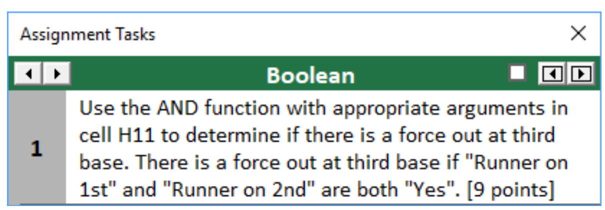
After the student enables macros, he or she is presented with the worksheets necessary to the assignment as well as a set of tools to manage both the completion and the submission of the assignment, as seen in Figure 3.

Figure 3. Assignment example



Detailed instructions on assignment requirements are included in the workbook and can be opened as a local HTML file (the Instruction Sheet) or presented one step at a time directly in Excel within a floating window (the Task Guide). When students have completed their work, the use the "Submit" tool to have their work graded. While students are working through assignment requirements, the workbook records every change they make to a cell as well as other activities such as adding worksheets and creating charts. The workbook also keeps track of when it is opened, each time the instruction sheet is shown, each time task guide (Figure 4) is advanced to show another task, and when the workbook is submitted.

Figure 4. Task guide



The data collected by this logging process provides a detailed history of *how* the student completed the assignment. Because each activity in the log is market with current time (down to the second), we can see not only what steps were taken but also how quickly they are accomplished. An example of data collected is seen here.

Figure 5. Data collected within Microsoft Excel

show task		
#	Time	Task
18	10/7/2017 10:19:18 AM	1.4 Fly ball complete
19	10/7/2017 10:19:22 AM	1.5 Not 2 outs

range modification						
#	Time	Worksheet	Address	Formula	Value	Step #
20	10/7/2017 10:19:45 AM	Boolean Functions	J11	=NOT(G11=2)	TRUE	1
21	10/7/2017 10:19:56 AM	Boolean Functions	J12:J40	=NOT(G12=2)	TRUE	1

Integrating the log of the student work with the log of watching videos and reading text content yields a rich picture of how the students engaged in the learning process. We can now see the extent to which they study instructional materials prior to working on the assignment, and we can tell when they pause in the completing the assignment. We can even tell whether they are reviewing the instructional content for the particular topic, or if they are viewing it for the first time.

One challenge in integrating these two different logs is that they are based on different clocks. The instructional log is based on the MyEducator servers, while the work log is based in the student's local computer. Therefore, during the submission process, the difference in the two clocks is recorded as is the network latency between the two machines so the logs can be synchronized with sub-second accuracy and the worklog from the student workbook is extracted and added to the server-side database

4. Analysis

Student data were coded into a string that included one letter per 10 percent finished, providing a human-friendly view into student strategies. The code allowed the researchers to visually inspect student strategies. Each letter represents 10 percent completion of different learning activity as follows:

Table 1. Code letters

r	reading instructional text
v	watching instructional video
p	completing tasks in practice assessments
a	completing tasks in primary (graded) assessments
t	viewing task instructions (task by task)
T	viewing task instructions (all tasks at once)

Using these letters, a code was created for each student within each chapter. As an exploratory step, we evaluated codes visually and made initial learning approach groups. Table 2 shows the breakdown of an example code.

Table 2. Example Code

`rrrrvvvrrppprrpptttaatataatatata`

`rrrr`	started by reading 40% of the chapter text
`vvv`	watched 30% of the video blocks
`rr`	read another 20% of the chapter text
`ppp`	completed 30% of tasks in practice assignment(s)
`ttt`	viewed 30% of the task-by-task instructions in the primary assignment
`aa`	completed 20% of the primary assignment
`tataat atata`	continued by alternating between instructions and primary assignment

The following are examples of additional codes for student study habits:

Table 3. Example codes

`rrrrrrrrt atataatat aaata`	student read and completed the assignment (no use of practice assignments or videos)
`tatattata tataatta`	student went straight to the assignment without reading the text or watching video (he or she may have already known the topic)
`vvvvvvrrr rrrrrrttt aattataat taata`	student watched video first, then read the text, and then completed the assignment
`taratarrt tatartaar rattta`	student started with the assignment and seems to have referenced the text when needed

With an initial understanding of student study habits from the codes, the source data were explored using a number of cluster analysis models. The final model included four groups, presented in Table 4.

The input data included percentages and/or counts of student activity in the five categories: reading, video watching, practice assignment actions, primary assignment actions, and task guide views. In addition to the percentages and counts, scores were calculated to show the order of activities and the level of overlap between the activities. For example, the codes `rrrvvv`, `vvvrrr`, `rvrvrv`, `vvrrvr` show the same completion of reading and video (30% each), but the order and overlap is different between them.

Table 4: Results of cluster analysis

Approach	Description	Proportion	Example codes from data
1. *Knowledge Confident*	Low Reading (20%). Low assignment/task description use. Moderate task assignment overlap.	26%	`rrrttttttTtataaaaaaat` `tttaataattaataatt`
2. *Confident Traditional*	Moderate to High Reading (65%). Low assign and task description use. High task assignment overlap.	29%	`rrrrrttttattatatataaaa` `rrrrTttttttttaaaaaataat`
3. *Less Careful Traditional*	High Reading (75%). High assignment use. Low task description use and task assignment overlap	18%	`rrrrrraaaaaaaaa` `rrrrrTaaaaaaaa`
4. *Diligent Traditional*	High Reading (80%). More reading on difficult topics less on possibly easier topics. High assignment and task description use. High assignment task overlap.	27%	`rrrrrrTttaataaaaaattttttt` `rrrrrTttatatTaaatattaat`

Figure 6. Student activity patterns

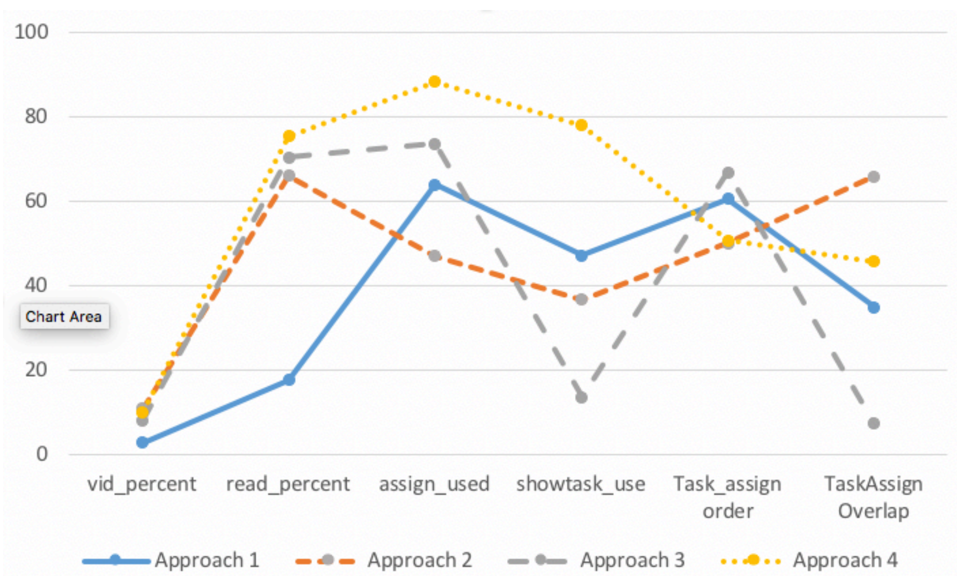


Table 5. Patterns and average scores by lesson

	Lesson 1: Excel Basics		Lesson 2: Functions and Formulas		Lesson 3: Logic and References		Overall	
	%	score	%	score	%	score	%	score
Approach 1: Knowledgeable Confident	29.3	97.6	33.3	97.8	33.3	98.1	32.0	97.8
Approach 2: Confident Traditional	33.5	96.3	35.6	97.5	35.6	97.2	34.9	97.0
Approach 3: Less Careful Traditional	16.8	94.2	15.8	94.0	15.3	92.6	16.0	93.6
Approach 4: Diligent Traditional	20.4	94.4	15.3	95.6	15.8	95.1	17.2	95.0

Figure 6 shows that strategies 1 and 2 are very similar to one another, with the primary difference being the amount of reading done. The students using these two approaches achieved higher scores than those using approaches 3 and 4.

Students tended to read and view videos first: less than 20% of the videos were watch by any of the groups. This finding confirms the same trend seen in an accounting course study [12], which also used the MyEducator platform.

Students in approaches 1, 2, and 4 tended to look at the task instructions before attempting the assignments. These groups also tended to work fairly linearly: they viewed task instructions and completed the assignment as separate and distinct activities (low task assignment overlap).

Learning strategy patterns varied the most in the amount of reading done and in how often students referred to the text while completing assignment tasks.

While all students shared a preference of reading the instructional materials over watching videos, they also tended to complete any reading or watching videos prior to attempting the assignment. Activity patterns varied primarily in the amount of reading done, as well as how often they went back to the assignment and task description as they completed each test your understanding problem. Student utilizing approach 2 were the only ones who tended to separate the viewing of the task and completing the assignment, while the others tended to go back and forth more between the two activities more often.

Many students (68%) switched approaches at least once while completing these three assignments. A third (37%) of the students maintained the same activity pattern for all three assignments. Approximately 48% of these students switch approaches between the Excel basics lesson and the functions and formulas lesson.

Readers should note that on lesson 3 (logic and references), students tended to read more of the text and get slightly lower scores than the other assignments. This is likely due to the increased difficulty of this assignment.

5. Conclusion

This study explored student study strategies in an introductory technology course. The technological tools in the course gave opportunity to collect detailed data regarding student reading, video watching, practice assignment activity, instruction viewing, and primary (graded) activity.

This exploratory study raises a number of questions that should be addressed in further experiments:

- Why don't students use video very much, especially when anecdotal evidence suggests that they do?
- How are different strategies correlated with student learning and/or performance?
- What strategies are most effective for specific personality types or learning styles?
- How course-specific and content-sensitive are the clusters found in this study?
- What are the primary indicators (and needed data) for identifying various strategies? How can these data be collected and analyzed efficiently in real time?
- How and why do students adjust their strategies as a course progresses?

The study provided insight into how students learn and enable instructors to better match their courses to student learning strategies. But beyond informing instructors, we hope the study contributes to the building of learning systems that *adapt in real time to student strategies*. Once strategies and indicators of those strategies are known, learning systems could be built to guide students on when and where to read, watch video, try practice assignments, and complete graded assignments—all tailored to the student as an individual learner.

6. References

[1] Nisbet, J., & Shucksmith, J. (2017). Learning strategies. Routledge.

[2] Oxford, R. (1990). Language learning strategies: What every teacher should know. Boston: Heinle and Heinle Publishers.

[3] Entwistle, N., & Ramsden, P. (2015). Understanding student learning (Routledge revivals). Routledge.

[4] Davies, R., Nyland, R., Bodily, R., Chapman, J., Jones, B., & Young, J. (2017). Designing technology-enabled instruction to utilize Learning Analytics. Tech Trends, 61(2). 155-161.

[5] Davies, R. (2002). Exploring the Meaning and Function of Learner Intent for Students taking Online University Courses. Published by Saarbrücken, Germany: VDM Publishing House Ltd., 2009.

[6] Dolmans, D. H., Loyens, S. M., Marcq, H., & Gijbels, D. (2016). Deep and surface learning in problem-based learning: a review of the literature. Advances in Health Sciences Education, 21(5), 1087-1112.

[7] Bonk, C. J., Lee, M. M., Kou, X., Xu, S., & Sheu, F. R. (2015). Understanding the self-directed online learning preferences, goals, achievements, and challenges of MIT OpenCourseWare subscribers. Journal of Educational Technology & Society, 18(2), 349.

[8] Ben-Eliyahu, A., & Bernacki, M. L. (2015). Addressing complexities in self-regulated learning: a focus on contextual factors, contingencies, and dynamic relations. Metacognition and Learning, 10(1), 1-13.

[9] Midgley, C. (Ed.). (2014). Goals, goal structures, and patterns of adaptive learning. Routledge.

[10] Kember, D., Biggs, J., & Leung, D. Y. (2004). Examining the multidimensionality of approaches to learning through the development of a revised version of the Learning Process Questionnaire. British Journal of Educational Psychology, 74(2), 261-279.

[11] Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. Metacognition and learning, 9(2), 161-185.

[12] James Stice, E. Kay Stice, and Conan Albrecht. "Study Choices by Introductory Accounting Students: Those Who Choose to Study by Reading Text Outperform Those Who Choose to Study by Watching Video Lectures". Proceedings of the AAA Conference: Western Region. 2016.